



УНИВЕРСИТЕТ ИТМО

Факультет ПИиКТ

Лабораторная работа №4 по Информатике  
Исследование протоколов, форматов обмена информацией  
и языков разметки документов  
Вариант 23

**Выполнил:**

Давааням Баясгалан

группа Р3111

**Преподаватель:**

Малышева Татьяна Алексеевна

г. Санкт-Петербург

2021 год

## Оглавление

Вариант .....	3
Исходный файл (XML).....	3
Обязательное задание .....	4
Код программы на Python .....	4
Результат программы.....	5
Дополнительное задание №1 .....	5
Код программы на Python .....	5
Результат программы.....	6
Дополнительное задание №2 .....	6
Код программы на Python .....	7
Результат программы.....	7
Дополнительное задание №3 .....	8
Код программы на Python .....	8
Результат программы.....	9
Вывод .....	9
Список литературы .....	9

## Вариант

23	XML	YAML	Четверг
----	-----	------	---------

## Исходный файл (XML)

```
<?xml version="1.0" encoding="UTF-8"?>
<timetable>
  <subject_1>
    <lesson>Математика (Практика)</lesson>
    <location>ул.Ломоносова д.9ь лит.М </location>
    <room>1512</room>
    <teacher>Холодова Светлана Евгеньевна</teacher>
    <time>8:20 - 9:50</time>
    <week>Все</week>
    <weekday>Четверг</weekday>
  </subject_1>
  <subject_2>
    <lesson>Математика (Практика)</lesson>
    <location>ул.Ломоносова д.9ь лит.М</location>
    <room>1329</room>
    <teacher>Холодова Светлана Евгеньевна</teacher>
    <time>11:40-13:10</time>
    <week>Все</week>
    <weekday>Четверг</weekday>
  </subject_2>
  <subject_3>
    <lesson>Математика (Лекция)</lesson>
    <location>ул.Ломоносова д.9ь лит.М</location>
    <room>1222</room>
    <teacher>Холодова Светлана Евгеньевна</teacher>
    <time>13:30-15:00</time>
    <week>Все</week>
    <weekday>Четверг</weekday>
  </subject_3>
</timetable>
```

## Обязательное задание

- Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.
- Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

### Код программы на Python

```
def parser():
    address_input = "D:\ITMO\1st Semester\Computer Science
(Basics)\lab4\Timetable_xml.xml"
    address_output = "D:\ITMO\1st Semester\Computer Science
(Basics)\lab4\Timetable_yaml.yml"
    input_file = open(address_input, 'r', encoding='utf-8')
    output_file = open(address_output, 'w', encoding='utf-8')
    lines = input_file.readlines()
    countTab = 0
    for line in lines:
        line = line.strip()
        if "<?" in line[:2]:
            continue
        elif "</" in line[:2]:
            countTab -= 1
        elif "<" in line[0]:
            start = line.index("<")
            end = line.index(">")
            cline = line[end+1:]
            if len(cline)<1:
                output_file.write(countTab * ' ' + line[start+1:end] + ":\n")
            )
            countTab += 1
        else:
            if cline.find("<") > -1:
                close_tag = cline.index("<")
                output_file.write(countTab * ' ' + line[start+1:end] + ":
" + cline[:close_tag] + "\n")
            else:
                output_file.write(countTab * ' ' + line[start+1:end] + ":
" + cline + "\n")
                countTab += 1
        else:
            if "<" in line:
                end = line.index("<")
                output_file.write(countTab * ' ' + line[:end] + '\n')
                countTab -= 1
            else:
                output_file.write(countTab * ' ' + line + '\n')
```

```
input_file.close()
output_file.close()
```

## Результат программы

```
timetable:
  subject_1:
    lesson: Математика (Практика)
    location: ул.Ломоносова д.9ь лит.М
    room: 1512
    teacher: Холодова Светлана Евгеньевна
    time: 8:20 - 9:50
    week: Все
    weekday: Четверг
  subject_2:
    lesson: Математика (Практика)
    location: ул.Ломоносова д.9ь лит.М
    room: 1329
    teacher: Холодова Светлана Евгеньевна
    time: 11:40-13:10
    week: Все
    weekday: Четверг
  subject_3:
    lesson: Математика (Лекция)
    location: ул.Ломоносова д.9ь лит.М
    room: 1222
    teacher: Холодова Светлана Евгеньевна
    time: 13:30-15:00
    week: Все
    weekday: Четверг
```

## Дополнительное задание №1

- Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
- Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
- Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.

## Код программы на Python

```
import xmlplain
def parser_with_library():
    address_input = "D:\ITMO\1st Semester\Computer Science
(Basics)\lab4\Timetable_xml.xml"
```

```
address_output = "D:\\ITM0\\1st Semester\\Computer Science
(Basics)\\lab4\\Timetable_with_library_yaml.yml"
input_file = open(address_input, 'r', encoding='utf-8')
output_file = open(address_output, 'w', encoding='utf-8')
root = xmlplain.xml_to_obj(input_file, strip_space=True, fold_dict=True)
xmlplain.obj_to_yaml(root, output_file)
input_file.close()
output_file.close()
```

## Результат программы

```
timetable:
  subject_1:
    lesson: Математика (Практика)
    location: 'ул.Ломоносова д.9ь лит.М '
    room: '1512'
    teacher: Холодова Светлана Евгеньевна
    time: 8:20 - 9:50
    week: Все
    weekday: Четверг
  subject_2:
    lesson: Математика (Практика)
    location: ул.Ломоносова д.9ь лит.М
    room: '1329'
    teacher: Холодова Светлана Евгеньевна
    time: 11:40-13:10
    week: Все
    weekday: Четверг
  subject_3:
    lesson: Математика (Лекция)
    location: ул.Ломоносова д.9ь лит.М
    room: '1222'
    teacher: Холодова Светлана Евгеньевна
    time: 13:30-15:00
    week: Все
    weekday: Четверг
```

## Дополнительное задание №2

- a) Переписать исходный код, добавив в него использование регулярных выражений.
- b) Сравнить полученные результаты и объяснить их сходство/различие.

## Код программы на Python

```
import re
def parser_with_re():
    address_input = "D:\\ITMO\\1st Semester\\Computer Science
(Basics)\\lab4\\Timetable_xml.xml"
    address_output = "D:\\ITMO\\1st Semester\\Computer Science
(Basics)\\lab4\\Timetable_with_re_yaml.yml"
    input_file = open(address_input, 'r', encoding='utf-8')
    output_file = open(address_output, 'w', encoding='utf-8')
    xml_lines = input_file.readlines()
    countTab = 0
    for xml_line in xml_lines:
        line = xml_line.strip()
        if re.search(r'<[a-я\\w\\d]*>[a-я\\W\\w\\d]*<\\/[a-я\\w\\d]*>', line, re.I):
            str = ' '.join(re.findall(r'<[a-я\\w\\d]*>', line))[1:-1]
            str = str + ": " + ' '.join(re.findall(r'>[a-
я\\W\\w\\d]*<\\/', line))[1:-2]
            output_file.write(countTab * ' ' + str + "\\n")
        elif re.search(r'<[a-я\\w\\d]*>', line, re.I):
            str_1 = ' '.join(re.findall(r'<[a-я\\w\\d]*>', line))[1:-1] + ':'
            output_file.write(countTab * ' ' + str_1 + "\\n")
            countTab += 1
        else:
            if re.search(r'^[a-я\\w\\d]+', line, re.I):
                str_2 = ' '.join(re.findall(r'[a-я\\w\\d\\D]+', line, re.I))
                print(str_2)
                output_file.write(countTab * ' ' + str_2 + '\\n')
                if re.search(r'<\\/[a-я\\w\\d]*>', line):
                    countTab -= 1
            elif re.search(r'<\\?', line):
                continue
            else:
                countTab -= 1
    input_file.close()
    output_file.close()
```

## Результат программы

```
timetable:
subject_1:
    lesson: Математика (Практика)
    location: ул.Ломоносова д.9ь лит.М
    room: 1512
    teacher: Холодова Светлана Евгеньевна
    time: 8:20 - 9:50
    week: Все
    weekday: Четверг
```

```
subject_2:
  lesson: Математика (Практика)
  location: ул.Ломоносова д.9ь лит.М
  room: 1329
  teacher: Холодова Светлана Евгеньевна
  time: 11:40-13:10
  week: Все
  weekday: Четверг
subject_3:
  lesson: Математика (Лекция)
  location: ул.Ломоносова д.9ь лит.М
  room: 1222
  teacher: Холодова Светлана Евгеньевна
  time: 13:30-15:00
  week: Все
  weekday: Четверг
```

## Дополнительное задание №3

- a) Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и программу из дополнительного задания №2, сравнить десятикратное время выполнения парсинга + конвертации в цикле.
- b) Проанализировать полученные результаты и объяснить их сходство/различие.

### Код программы на Python

```
import time
parser_time = time.time()
for i in range(10):
    parser()
parser_time = time.time() - parser_time

parser_with_library_time = time.time()
for i in range(10):
    parser_with_library()
parser_with_library_time = time.time() - parser_with_library_time

parser_with_re_time = time.time()
for i in range(10):
    parser_with_re()
parser_with_re_time = time.time() - parser_with_re_time

print('Parser time: ' + str(parser_time))
print('Parser time with library: ' + str(parser_with_library_time))
print('Parser time with re: ' + str(parser_with_re_time))
```



## Результат программы

```
C:\Users\HP>C:/Users/HP/AppData/Local/Programs/Python/P
Parser time: 0.00797891616821289
Parser time with library: 0.04215192794799805
Parser time with re: 0.016954660415649414

C:\Users\HP>C:/Users/HP/AppData/Local/Programs/Python/P
Parser time: 0.0057752132415771484
Parser time with library: 0.034307003021240234
Parser time with re: 0.009446859359741211

C:\Users\HP>C:/Users/HP/AppData/Local/Programs/Python/P
Parser time: 0.004985809326171875
Parser time with library: 0.043219804763793945
Parser time with re: 0.010935544967651367
```

## Вывод

При выполнении лабораторной работы я ознакомился с XML и YAML и научил их синтаксис. В языке Python я научил использование готовых библиотек и заметил разницу времени выполнения между готовым библиотекой и сам написанным конвертором.

## Список литературы

<https://habr.com/ru/post/309242/>

[https://ru.wikipedia.org/wiki/Форма\\_Бэкуса\\_—\\_Наура](https://ru.wikipedia.org/wiki/Форма_Бэкуса_—_Наура)